Locus-specific gene expression pattern suggests a unique propagation strategy for a giant algal virus

JOURNAL OF VIROLOGY, Aug. 2006, p. 7699–7705 Vol. 80, No. 15
0022-538X/06/$08.00+0 doi:10.1128/JVI.00491-06

Locus-Specific Gene Expression Pattern Suggests a Unique Propagation Strategy for a Giant Algal Virus†

Michael J. Allen,[1] Thorsten Forster,[2] Declan C. Schroeder,[3] Matthew Hall,[3] Douglas Roy,[2] Peter Ghazal,[2] and William H. Wilson*[1]

Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, United Kingdom[1]; Scottish Centre for Genomic Technology and Informatics, University of Edinburgh, Chancellor's Building, College of Medicine, 49 Little France Crescent, Edinburgh, EH16 4SB, United Kingdom[2]; and Marine Biological Association, Citadel Hill, Plymouth, PL1 2PB, United Kingdom[3]

Received 9 March 2006/Accepted 10 May 2006

***Emiliania huxleyi* virus strain 86 is the largest algal virus sequenced to date and is unique among the *Phycodnaviridae* since its genome is predicted to contain six RNA polymerase subunit genes. We have used a virus microarray to profile the temporal transcription strategy of this unusual virus during infection. There are two distinct transcription phases to the infection process. The primary phase is dominated by a group of coding sequences (CDSs) expressed by 1 h postinfection that are localized to a subregion of the genome. The CDS of the primary group have no database homologues, and each is associated with a unique promoter element. The remainder of the CDSs are expressed in a secondary phase between 2 and 4 hours postinfection. Compartmentalized transcription of the two distinctive phases is discussed. We hypothesize that immediately after infection the nucleic acid of the virus targets the host nucleus, where primary-phase genes are transcribed by host RNA polymerase which recognizes the viral promoter. Secondary-phase transcription may then be conducted in the cytoplasm.**

Emiliania huxleyi virus strain 86, EhV-86, is the type species of the genus *Coccolithovirus* within the family of algal viruses, the *Phycodnaviridae* (14). EhV-86 is a large, double-stranded DNA virus that infects the marine coccolithophorid *Emiliania huxleyi* (13), a unicellular alga known for its elegant calcium carbonate scales (coccoliths) that are produced intracellularly and sequestered over its outer cell surface (21). *E. huxleyi* is perhaps best known for its immense coastal and open-ocean blooms at temperate latitudes and is a key species for current studies on global biogeochemical cycles and climate modeling (6, 19, 20).

We have previously sequenced the 407,339-bp genome of EhV-86 and predicted that it encodes 472 protein coding sequences (CDSs) (22). This is the largest *Phycodnaviridae* genome sequenced to date. The functions of the 472 CDSs are largely unknown: only 66 have been annotated with functional product predictions on the basis of sequence similarity or protein domain matches (22). Genes of known function encode DNA polymerase subunits, RNA polymerase subunits, sphingolipid biosynthesis enzymes, and eight proteases. The *Phycodnaviridae* family was previously thought to lack RNA polymerase; thus, the detection of six RNA polymerase subunits in EhV-86 suggested a lifestyle distinct from those of other *Phycodnaviridae* (1–3, 22). A lytic–phase transcriptional profile generated from *E. huxleyi* infected by EhV–86, 33 h postinfection, detected the presence of 65% of the predicted virus CDS transcripts (22).

In many well-studied biological systems, viral transcripts are commonly assigned into a kinetic class by their dependence on protein synthesis or viral DNA replication (5). This is usually achieved by monitoring the sensitivity of virus gene transcript levels upon exposure to inhibitory drugs (5). Through the inhibition of translation or DNA replication with chemicals such as cycloheximide and phosphonoacetic acid, transcripts can be assigned into kinetic classes such as immediate-early (α), early (β), early-late ($\gamma 1$), and late ($\gamma 2$). Kinetic stratification strongly correlates with functional classification. Classically, immediate-early proteins perform regulatory functions; early proteins, metabolic roles; and late proteins, structural and morphogenic roles. The culturing conditions needed to sustain optimum growth in tandem with the number of cells needed to extract total RNA in useable quantities (experiments typically involve tens of liters of culture medium) limit what can be done presently with the *E. huxleyi* experimental system. The application of inhibiting drugs to *E. huxleyi* cultures is both impractical and costly. Furthermore, the molecular biology of *E. huxleyi* is almost as poorly understood as the molecular biology of its virus, EhV-86. The use of drugs may have unknown affects on both host and virus systems. To build on previous genome sequencing work and to begin the task of characterizing all the EhV-86 CDSs, the majority of which are of unknown function, an expression profile of the first 4 h of infection was determined. We know from experience that mature virions are released from infected cells 4 h postinfection (data not shown); hence, this initial period was profiled in detail. By exploiting a synchronous infection and sampling early on, we avoided the need for inhibitor drugs and profiled a natural infection cycle.

* Corresponding author. Mailing address: Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, United Kingdom. Phone: 44 1752 633100. Fax: 44 01752 633101. E-mail: whw@pml.ac.uk.

† Supplemental material for this article may be found at http://jvi.asm.org/.

MATERIALS AND METHODS

Growth conditions and RNA extraction. Experimental methodology is described in depth in the MIAME compliant database entry E-MAXD-8, available at www.arrayexpress.com. Briefly, exponentially growing *Emiliania huxleyi* (1 ×

10^6 cells ml^{-1}) was infected with EhV-86 (fresh lysate, 5×10^8 virions ml^{-1}) in f/2 medium at 15°C in a Sanyo MLR-350 incubator with a ratio of 16:8 h of light-dark illumination.

Total RNA was extracted 0, 1, 2, and 4 h postinfection, using a QIAGEN midiprep kit, and used for microarray analysis as described previously (22). Briefly, cell cultures (150 ml) were filtered (0.45-μm Supor-200 membrane filter; Pall Corporation), resuspended in 2 ml phosphate-buffered saline solution, pelleted by centrifugation (20,000 × g, 1 min), resuspensed in 2 ml RNAlater (QIAGEN), and stored at −20°C until processed with the QIAGEN midiprep kit.

As a control, virus particles were checked for the presence of RNA. The virus particles were purified and concentrated as described previously (14). Briefly, virus lysate (5 liters) was filtered (0.45-μm Supor-200 membrane filter; Pall Corporation), concentrated by a tangential flow ultrafiltration unit with a molecular weight cutoff of 50,000 (Vivaflow 50; Sartorius) to a final volume of 20 ml. Aliquots (3.5 ml) of the concentrate were adjusted with CsCl to densities of 1.1, 1.2, 1.3, and 1.4, and gradients from 1.1 and 1.4 were created by ultracentrifugation at 25,000 rpm at 22°C for 2 h in a SW40 Ti Beckman rotor. Virus bands were removed with a syringe and dialyzed against 4 × 1-liter volumes of 10 mM Tris-HCl, pH 7, at 4°C. To remove any contaminating RNA associated with the outside of the virus particles, the concentrate was incubated with RNAseA (final concentration, 50 μg ml^{-1}) for 1 h at 37°C and treated with Proteinase K (final concentration, 50 μg ml^{-1}) and sodium dodecyl sulfate (SDS) (final concentration, 0.5%), and the samples were incubated at 60°C for 1 h. Samples were then treated to extract total RNA as described previously. Purified total RNA was then treated with RNase free DNase (QIAGEN) in a 50 μl volume. Following DNase treatment, RNA was cleaned and purified, using the Roche Target Purification kit (Roche), and eluted in 50 μl RNase-free water and stored at −80°C. RNA quantity and quality were assessed using the Agilent Bioanalyzer 2100 system. All RNA samples were checked for virus genomic DNA contamination by PCR, using virus major-capsid protein gene primers (14). All samples tested negative for viral genomic DNA contamination.

Fluorescent labeling of cRNA. Random amplification of the entire mRNA population was achieved using the Microarray Target Amplification kit (Roche). This requires the use of a special primer containing a random sequence with no significant homology to any known sequences in public databases (Target Amplification Sequence [TAS]) in addition to a T7 promoter and oligo(dT) sequences for first-strand cDNA synthesis. The TAS region generates a 3′ anchor on the cDNA for subsequent PCR amplification with a TAS-PCR primer. First- and second-strand cDNA synthesis was performed from 200 ng of template RNA and 10 pg of spike mRNA (Stratagene), and the product was purified using the Target Purification kit (Roche). Purified cDNA was then randomly amplified by 24 cycles of PCR with TAS primers. PCR products were purified using the Microarray Target Purification kit (Roche) and then concentrated using YM30 Microcon concentration columns (Millipore). The resulting PCR-amplified cDNA was then transcribed into Cy3 fluorescently labeled cRNA, using the T7 Microarray RNA Target Synthesis kit (Roche). Labeled cRNA was then purified using the Microarray Target Purification kit. The EhV-86 microarray has been described previously (22). Briefly, 75-mer oligonucleotides are spotted in triplicate for 425 of the 472 predicted CDSs of EhV-86. For hybridization to microarrays, 7.5 μl 20 × SSC (Sigma), 1.0 μl 10% SDS, and the labeled cRNA sample were combined in a total volume of 50 μl. After being incubated at 100°C for 2 min (lid temperature, 110°C), the samples were cooled (room temperature, 5 min) and the volume was checked (and made up to 50 μl when necessary) prior to loading the samples onto the microarray slide covered by a LifterSlip (Erie Scientific Company). Microarray hybridization was performed in a microarray hybrid chamber (Camlab) at 65°C for 18 h. The microarray slides were given three posthybridization washes. The first wash was in 50 ml of 1 × SSC–0.1% SDS for 5 min with constant agitation, followed by a second wash for a further 5 min (with constant agitation) in 50 ml of 0.1 × SSC–0.1% SDS. A third wash was performed by plunging the slides 20 times into 50 ml of 0.1 × SSC. The microarray slides were immediately centrifuged to remove residual liquid (200 g, 1 min) and stored in the dark prior to being scanned. Three biologically independent samples for each of the four main study conditions (uninfected *Emiliania huxleyi*, postinfection at 0 h, 1 h, 2 h, and 4 h) were hybridized to one array each. For validation purposes, three further arrays were hybridized with RNA extracted from purified EhV-86 virion particles, and one array was hybridized with just the Stratagene spike RNA.

Microarray data processing and analysis. (i) Scanning and image processing. Hybridized arrays were scanned with an Affymetrix 418 array scanner. Six scans at incremental settings (gains of 0, 5, 10, 15, 20, and 25) were performed for each array in order to determine and select the optimal scan setting, producing a high dynamic signal range without saturation (9). Images from all scans were quantified using GMS Scanner software, version 1.51.0.42, and ImaGene, version 5.6.1.

All subsequent data processing and analysis steps were carried out with the R version 2.1.1 statistical programming environment (www.r-project.org) and Bioconductor version 1.6 microarray modules (www.bioconductor.org).

(ii) Background correction. Assessment of global noise and signal distributions showed evidence for a background noise gradient across some arrays. The effect was only noticeable at very low fluorescence levels and had no corresponding visible effect on signal levels. A log-linear interpolation method (Edwards) was used to correct lower-intensity values for any background carryover effects (8).

(iii) Normalization. A targeted viral array is likely to violate the assumptions that are made before applying global normalization methods; in this case it can be expected that 10% or more of the probes on the array will change expression levels between different biological conditions or time points. In anticipation of this, a large set of positive-control probes were included on the array (10 Stratagene alien probes, printed in triplicate on each of the 16 subgrids). The chosen normalization method was scaling of subset (n = 480 probes) medians to a common reference value, after an initial $\log_2$ transformation of all data values.

(iv) Expression determination. Expressed and/or not expressed calls for each gene and time point were generated on the basis of a detection threshold determined by the ninetieth percentile value of negative-control probes. A probe with an intensity value above this threshold was considered to be "on."

(v) Differential expression and hypothesis testing. Prior to further analysis, a nonspecific filter was applied to remove biologically irrelevant genes (control probes). This process reduces problems with statistical testing on multiple variables simultaneously. Of a total of 2,496 probes, 1,440 remained in the analysis set. The threshold for the filtering was in this case based on the slightly less stringent eightieth percentile of negative-control probe intensities. Differential expression was determined in two ways: (i) from an uninfected baseline sample to each of the postinfection time points and (ii) between postinfection time points. A statistical test with relative robustness for small sample sizes (simple linear model enhanced by empirical Bayes) was used to test the null hypothesis of nondifferential expression for each individual gene. Simultaneous testing on large numbers of variables (genes) leads to an increased number of potentially false-positive results; a Benjamini and Hochberg P-value correction was therefore applied.

(vi) Explorative analysis. Prior to explorative analysis, a further nonspecific filter was applied to remove genes which were not "on" in at least 3 out of 18 arrays, reducing the number of probes to 1,288. Absolute expression data were grouped and visualized using a hierarchical biclustering algorithm on expression profiles of genes and samples, combined with a heat map. The clustering parameters were "1 minus correlation" as the distance measure and "average" as the linkage method. The gene clusters were standardized to have means and standard deviations of 0 and 1, respectively, across arrays.

Microarray data accession number. Microarray data (including microarray design, hybridization, and analysis) were stored and curated in maxdLoad2 prior to submission in MAGE/ML format to the EBI Array Express database (http://www.ebi.ac.uk/arrayexpress) under accession number E-MAXD-8 (10). This data is also available at the EGTDC data catalog (http://envgen.nox.ac.uk/) under accession number egcat:000010.

RESULTS AND DISCUSSION

Characterization of temporal expression profile. By random amplification of mRNA extracted from infected cells and using microarrays, the expression profile of the EhV-86 infection of *E. huxleyi* was determined over the first 4 h of infection (see Table S1 in the supplemental material). RNA samples were extracted at 0, 1, 2, and 4 h postinfection. CDSs were categorized into group T1, T2, or T4, based upon whether their expression was first detected at 1, 2, or 4 h postinfection, respectively (Fig. 1). Thirty-nine CDSs were assigned to group T1, 194 CDSs to group T2, and 71 CDSs to group T4. Transcripts for a further 115 were not detected, including 51 that have previously been detected in the 33-h-postinfection lytic-phase transcriptional profile (22), suggesting that the virus infection cycle may not be fully completed at 4 h postinfection. However, the transcript for the major capsid protein gene, ehv085, was detected at 4 h postinfection, suggesting that, at

T1 Transcripts	T2 Transcripts	T4 Transcripts
218, 219, 220, 233, 235, 242, 247, 249, 251, 252, 263, 265, 266, 271, 274, 279, 280*, 289, 290, 293, 294, 295, 296, 297, 298, 300, 303, 304, 305, 308, 311, 314*, 315, 316*, 320, 322, 323, 348, 349*	001, 006, 007, 008*, 009, 011, 014, 017, 018, 019, 020, 022, 024, 026, 028, 030, 031, 032, 033, 035, 041, 050, 051, 052, 053, 054, 057, 058, 061, 063, 064, 065, 066, 068, 073, 075, 076, 077, 079, 082, 084, 087, 090, 091, 094, 095, 103, 104, 105, 106, 107, 108, 110, 113, 115, 120, 121, 122, 123, 124, 127, 129, 130, 132, 133, 134, 136, 138, 139, 141, 142, 144, 148, 149, 151, 153, 155, 156, 158, 162, 163, 169, 174, 176, 179, 183, 184, 185A,186, 187, 188, 191, 192, 197, 198, 204, 208, 209, 210, 211, 212, 215, 216, 217, 221, 223, 226, 227, 230, 237, 241, 244, 245, 254, 256, 260, 270, 269A, 275, 281, 284, 286, 288, 302, 307, 312, 313, 328, 331, 334, 335, 336, 338, 339, 342, 344, 346, 350, 353, 355, 357, 359, 360, 367, 368, 369, 370, 372, 373, 378, 379, 380, 392, 395, 397, 399*, 401, 404, 405, 406, 414, 417, 419, 420, 421, 422, 423, 424, 425, 428, 429, 431*, 434, 435, 436, 437, 438, 439, 440, 445, 446, 447, 448, 449, 451, 453, 456, 459, 461, 464, 465, 466, 467, 468	002, 004, 005, 015, 016, 023, 029, 036, 037, 038, 047, 059, 062, 070, 071, 074, 083, 085, 086, 097, 098, 099, 100, 101, 112, 118, 119, 140, 143, 146, 147, 161, 165, 167, 177, 181, 185, 190, 201, 205, 214, 243, 264, 282, 291, 292, 330, 332, 351, 352, 354, 382*, 384, 385, 387, 388, 391, 393, 394*, 403, 407, 410, 411, 418, 426, 430, 452, 454, 455, 462, 463

FIG. 1. Grouping of EhV-86 CDSs. EhV-86 CDSs are listed at the experimental time point at which they first show statistically significant ($P < 0.01$) differential expression compared to that of baseline uninfected samples (T1, T2, and T4 indicate 1, 2, and 4 h postinfection, respectively). CDSs are included in this table only if they match two criteria. The first criterion requires CDSs to have signal intensity values above the detection threshold for at least two out of three print replicates (asterisks indicate two probe replicates in agreement). The detection threshold is defined nonparametrically as the ninetieth percentile of human negative-control probes. The second criterion requires the differential expression compared to that of uninfected samples to be statistically significant for all three print replicates of a CDS.

this point, the virus is in the latter stages of the infection process.

Early transcripts. During the preliminary annotation of the EhV-86 genome, CDSs were numbered sequentially, starting with ehv001 (bases 276 to 1,022) through to ehv468 (bases 406,039 to 406,896). The subsequent addition of any extra CDSs was achieved by naming according to the nearest CDS and adding an "A" suffix (e.g., ehv185A). Hence, the CDS number indicates the genomic location of a CDS relative to the neighboring CDSs. The 39 T1 virus transcripts are all localized to a specific region of the EhV-86 genome and have CDS numbers ranging between 218 and 366 (Fig. 1). These CDSs are found in a 104-kb section of the genome that has previously been identified as containing unique putative promoter elements known as family A repeats (3). The 151 CDSs in this region have few database homologues, and their origin and function are completely unknown (2, 3). Family A repeats are characterized by the presence of the conserved nonamer GT TCCC(T/C)AA that is found directly upstream of the predicted start of translation methionine codon (ATG) for 87 of the CDSs in this region. Expression of 39 CDSs from this region (P values < 0.01) at only 1 h postinfection, in combination with no expression from CDSs outside of this region, is a significant finding (Fig. 1; see also Table S2 in the supplemental material for P values).

Genes in the 104-kb central region not associated with the putative promoter show a different absolute expression profile to those associated with the putative promoter (Fig. 2). Two

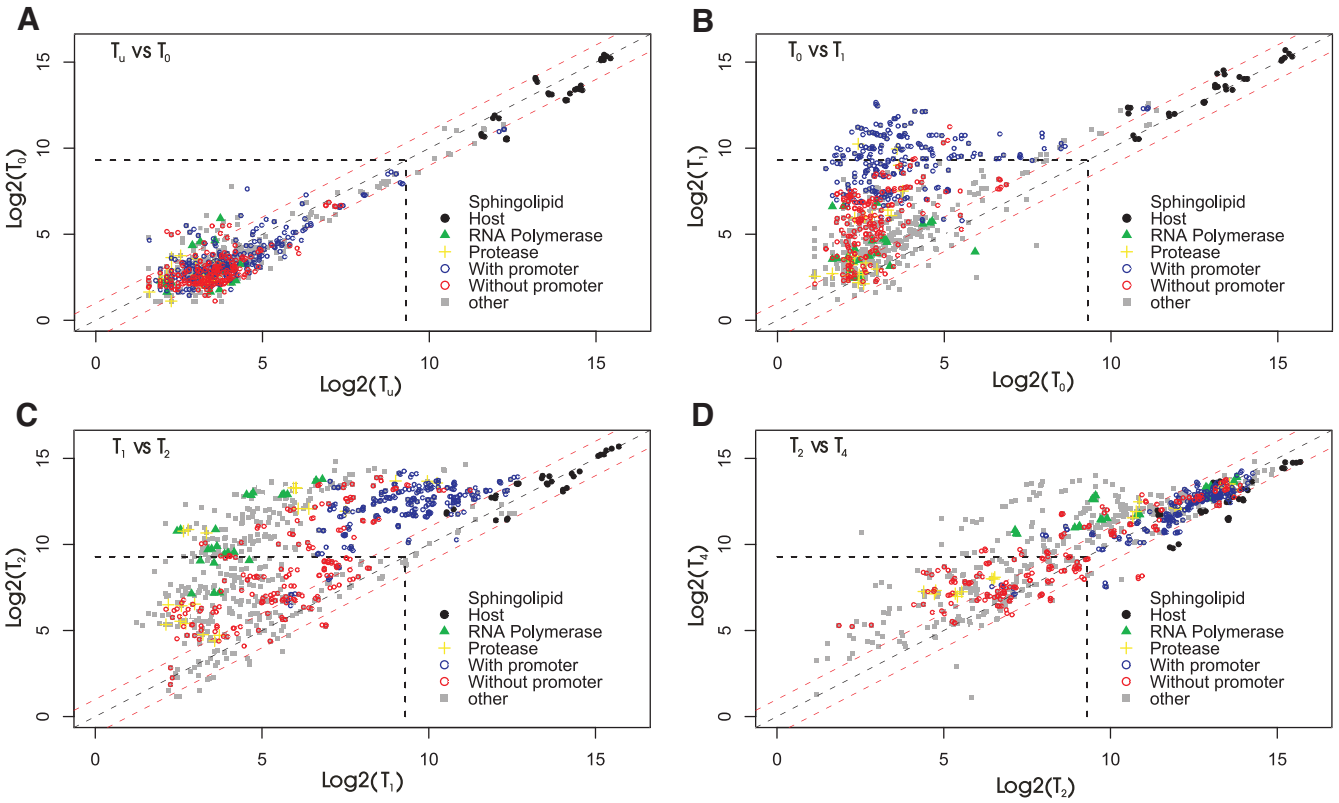

FIG. 2. Scatterplots of absolute CDS expression for uninfected versus each of the infected samples. The center diagonal lines represent "no difference in expression," and the two diagonal lines parallel to the center line represent the threshold for twofold-up and twofold-down regulation. The rectangular box outlined with a dashed line indicates the detection threshold for each of the two biological conditions; i.e., probes inside this box are not expressed or not measurable in uninfected and infected samples. The three replicate values for each CDS are shown.

apparently non-promoter-associated CDSs can be seen to cluster with the promoter-associated CDSs; these correspond to ehv349, which encodes a putative protease. This CDS is associated with the promoter element but was grouped with the proteases for the purpose of this analysis. The three replicate probes for ehv247 (which is not directly associated with the promoter) can be seen to cluster with the promoter-associated CDSs. It is likely that this CDS is cotranscribed with the immediately adjacent ehv248, which does have the promoter.

Those associated with the putative promoter have the highest expression levels at 1 h postinfection, whereas the CDSs in this region not associated with the putative promoter show similar expression profiles to those from the remainder of the genome. This early expression pattern provides important new evidence that family A repeats do function as promoter elements. Indeed, it has previously been suggested that family A repeats may function as immediate-early promoters (2, 3, 22). Expression driven by these promoter elements is likely to play an integral role during the early stages of the virus infection process. However, the lack of database homologues for the CDSs putatively driven by these promoters reveals no obvious insights into the function of this group of genes. Their early expression suggests that they could perform a regulatory role akin to the immediate-early genes identified in other viruses (4, 7, 16). Future functional analysis of this region will provide vital clues to the life cycle of this virus and the unique role of this 104-kb region.

Unexpectedly, expression of RNA polymerase subunit genes (ehv064, ehv108, ehv167, ehv399, and ehv434) does not appear to occur until later on in the infection process (transcripts for RNA polymerase subunits are first detected at 2 h postinfection) (Fig. 1), suggesting that the host RNA polymerase is responsible for expression of T1 genes or that the virus packages its own RNA polymerase into the virion.

Differential expression. An explorative analysis using biclustering and heat map visualization of absolute expression values (Fig. 3) reveals that the uninfected transcriptional profile (T_u) is, as would be expected, very similar to the infected profile immediately after infection (T_0), with only a handful of host genes expressed on the array. Host genes display a general trend of down regulation (although this down regulation is not statistically significant) which accounts for the minor differences between the T_u and T_0 profiles. These changes are likely to have been caused by the unavoidable delay between sampling cells and freezing of the RNA profile (approximately 5 min), indicating that the virus may have an effect on the host cell within minutes of infection. Two-hour and 4 h postinfection transcriptional profiles (T_2 and T_4, respectively) are very similar and cluster tightly together with transcripts for the majority of EhV-86 CDSs expressed (Fig. 1 to 3). The 1 h postinfection transcriptional profile (T_1) fails to associate closely with either of these two clusters but appears closer to the T_u and T_0 profiles than to T_2 and T_4 (Fig. 3).

The similarity in profiles between T_u and T_0 and between T_2

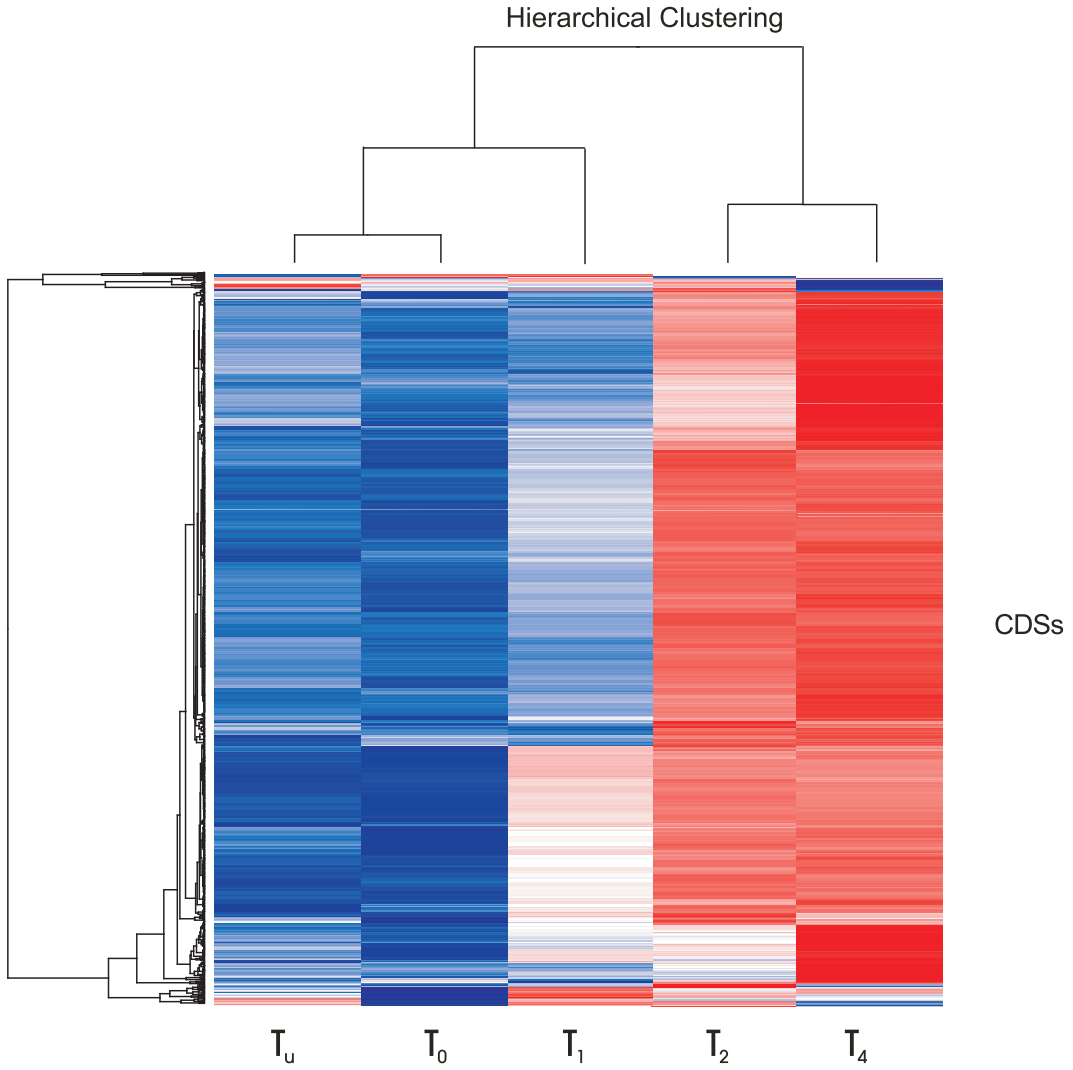

FIG. 3. Biclustering of CDS and sample expression profiles in combination with heat map visualization. Each column represents the averaged $\log_2$ expression data for three independent samples for a biological condition. Each row in the heat map represents one CDS color coded for expression level from low (blue) to medium (white) to high (red). Branches on the top and side of the heat map show the identified clusters in the sample, i.e., samples with the most-similar expression profiles across all genes and genes with similar expression profiles across all samples. As expected, results for uninfected (T_u) samples and samples at 0 h postinfection (T_0) are very closely related. The later time points T_2 and T_4 are also closely related. T_1 is the most interesting, showing a cluster of CDSs which go from "off" in the T_u and T_0 samples to "on."

and T_4 is represented by the linear distribution observed in Fig. 2A and D. Viral transcripts are expressed at increasing levels for the first 2 h of infection (Fig. 2B and C), after which the expression levels of the majority of CDSs are maintained at a more consistent level.

Virion message. When a target was labeled from a total RNA extraction performed on purified EhV-86 virions as a control, the only transcript detected above threshold levels on the microarray was that of ehv315 (P value < 0.01) (data not shown). CDS ehv315 is a proline-rich putative membrane protein of unknown function (22). Proline-rich proteins have previously been implicated in calcium binding (22), which is of particular interest since the host, *E. huxleyi*, is well known for the sequestration of calcium carbonate onto its surface (in the form of coccoliths) during active growth (21). Whether the protein EHV315 plays a role in disrupting this pathway is yet to be determined but clearly warrants further investigation.

CDSs not expressed. Many of the EhV-86 CDSs where no transcript was detected at 4 h postinfection have been assigned a putative function (22) (Fig. 4). Noteworthy is the lack of detection of transcripts for five of the eight proteases predicted to be encoded in the EhV-86 genome (ehv021, ehv109, ehv160, ehv349, and ehv361) (22). Virally encoded proteases have previously been shown to be involved in the maturation of infectious virions, suggesting further that at 4 h postinfection the full virus replication cycle may not yet have been completed (15). At 4 h postinfection, the cell has only just started releasing virus particles and we know it continues to release viruses for up to 2 days while remaining intact (data not shown). This leaves scope for a range of new transcripts to be up-regulated

Transcripts not detected	Expression not confirmed	Not tested
003, 010, **013**, **021**, **025**, **027**, 034, 039, **040**, 042, 043, **044**, 046, **048**, 049, 056, 060, 067, 069, **072**, 078, **080**, **088**, 089, 092, 096, 102, **109**, 116, 117, 128, 131, **135**, 137, 145, **150**, **152**, **154**, 159, **160**, **164**, 166, 168, **170**, 171, **172**, **173**, **175**, 178, **182**, 195, 196, **199**, **200**, **206**, **210A**, **213**, 222, **228**, **229**, **231**, **232**, **234**, **246**, **250**, **253**, 259, **267**, **268**, **272A**, **277**, **278**, **301**, **306**, **308A**, 310, 319, 321, 324, 325, 326, 333, **337**, 340, 341, 345, 347, 356, 358, 361, 363, 365, 366, 371, 374, 375, 383, 386, 389, 390, 396, 398, 400, 402, **408**, 409, **412**, **413**, 415, **427**, 433, 441, **442**, 443, **450**	045, 055, 111, 189, 207, 272, 299, 329, 376, 432, 444,	012, 081, 093, 114, 125, 157, 180, 193, 194, 202, 203, 224, 225, 236, 238, 239, 240, 248, 255, 257, 258, 261, 262, 269, 273, 276, 283, 285, 287, 309, 317, 318, 327, 343, 362, 364, 377, 381, 416, 457, 458, 460

FIG. 4. EhV-86 CDSs for which expression of the transcript could not be detected above threshold levels, confirmed, or tested. CDSs that have previously been detected in a 33-h postinfection lytic-phase transcriptional profile are in bold (22). A transcript is defined as "not detected" if at least two out of three print replicates for this gene have signal intensity values at or below the ninetieth percentile for the signal intensity value of human negative-control probes.

at specific time points after this initial 4-h period. Moreover, the EhV-86 genome also encodes key components of sphingolipid biosynthesis, a pathway that leads to the production of ceramide (22), and these genes are expressed at 2 h postinfection (Fig. 1). It is known that there is a connection between protease activation and ceramide-induced apoptosis (17). A virus-controlled apoptotic mechanism is unlikely to induce cell death in the early stages of infection; consequently, protease expression will most likely be up-regulated closer to the onset of apoptosis and cell disintegration, up to 2 days later.

Other transcripts notable for their lack of detectable transcription include those coding for a putative lectin protein (ehv060), DNA-binding proteins (ehv072 and ehv152), phosphate permease (ehv117), thioredoxin (ehv358), protein kinase (ehv402), esterase (ehv363), fatty acid desaturase (ehv415), and a topoisomerase (ehv444). Indeed, for a number of CDSs transcripts have not been detected in the previous lytic-phase transcriptional profile (generated 33 h postinfection) or during the work described here (Fig. 4) (22). Four of these are "core genes" (ehv072, ehv128, ehv166, and ehv444) previously identified as being conserved in the nuclear-cytoplasmic large DNA virus (NCLDV) family. Their high degree of conservation in this family would suggest a role crucial to the virus, and with such a role they would be expected to be expressed. The lack of detection of these transcripts could be caused by their lack of expression or their expression at levels too low to be detected. Future work involving real-time PCR may determine whether these CDSs are expressed and at what levels.

Closing discussion. The infection cycle of EhV-86 can be divided into two broad stages: a primary stage in which a distinctive subgroup of localized CDSs associated with a putative promoter element are transcribed and a secondary stage during which CDSs are transcribed regardless of their genomic location. The function of the primary stage is difficult to ascertain, since the vast majority of the CDSs expressed have little or no database homologues. CDSs from this region have been shown to have some of the highest levels of expression during the infection process (22), presumably due to their early and then constant high levels of expression, suggesting that they are of vital importance to the infection strategy.

The presence of RNA polymerase genes in the EhV-86 genome implies that the virus has the capacity to transcribe its own genes from within the cytoplasm during infection. It is not implausible to suggest that the biphasic stages of expression that occur during EhV-86 infection of *E. huxleyi* are compartmentalized. There is no evidence that transcription of virus RNA polymerase genes occurs until at least 1 h into the infection; therefore, there are two main possibilities that could account for the expression of the primary genes: (i) a functional viral RNA polymerase that recognizes family A promoters is packaged into the mature virion and causes expression of primary CDSs in the cytoplasm, or (ii) following infection, viral DNA targets the host nucleus where host RNA polymerase recognizes family A promoters, leading to transcription of primary genes. Targeting of virus genomic DNA to the nucleus has previously been suggested to occur during the infection of

Chlorella by PBCV-1 (18). The change from nuclear to cytoplasmic intracellular compartments (or vice versa) could account for the distinctive change in transcriptional profile observed in these experiments. If promoters similar to the family A repeats are identified during the genomic sequencing of *E. huxleyi* (currently under way), the second option above should certainly gain greater credence. If this is the case, subsequent breakdown of host genomic DNA (ehv041 encodes an endonuclease and is expressed between 1 and 2 h postinfection) and the nuclear envelope (a viral protease, encoded by ehv349, is expressed during the first hour of infection) could initiate the secondary stage of infection, where the remaining virus genes are expressed.

The presence of RNA polymerase in the EhV-86 genome, so far unique among the *Phycodnaviridae*, suggested a unique replication cycle for this giant virus. The virally encoded RNA polymerase may be intrinsically linked to the virus-encoded sphingolipid biosynthesis pathway (22). Sphingolipids are well documented as playing a crucial role in controlling cell death (12). If degradation of host DNA does occur during infection, then the ability to make new host-encoded RNA polymerase would be lost. Furthermore, if cell death is delayed extensively by the manipulation of sphingolipid biosynthesis, the active production of a virus-encoded RNA polymerase during the prolonged infection would prove vital to the replication strategy of the virus. Consequently, this could account for the retention of RNA polymerase function in coccolithoviruses (2).

The NCLDV group is composed of the *Asfariviridae*, *Poxviridae*, *Mimiviridae*, *Iridoviridae*, and *Phycodnaviridae* families (2). These diverse families are likely to have shared a common ancestor which was likely to have had both nuclear and cytoplasmic phases in its life cycle (11). Lineage-specific gene loss and gain within the NCLDV families is thought to contribute to the highly diverse characteristics of present-day forms. Poxviruses, asfarviruses, and iridoviruses encode their own transcription and replication machinery and undergo their replication cycles entirely in the cytoplasm (poxviruses) or start it in the nucleus and complete it in the cytoplasm (asfariviruses and iridoviruses) (11). Prior to the sequencing of EhV-86, the phycodnaviruses were thought to be characterized by their nuclear-dependent replication cycles. This interesting coccolithovirus appears to have a replication cycle more similar to that of the ancestral virus, a cycle distinct from all currently known *Phycodnaviridae*.

ACKNOWLEDGMENTS

This research was supported by grants awarded to W.H.W. from the Natural Environment Research Council (NERC) Environmental Genomics thematic program (NE/A509332/1) and from Marine Genomics Europe through framework program FP6 of the European Commission and to P.G. from the Biotechnology and Biological Sciences Research Council (BBSRC), the Scottish Higher Education Funding Council (SHEFC), and the EU.

We acknowledge help with microarray data submission from the NERC Environmental Bioinformatics Centre, Centre for Ecology and Hydrology, Oxford, United Kingdom.

M.J.A. designed and performed the research, analyzed the data, and wrote the manuscript. W.H.W. conceived, designed, and coordinated the research. T.F. analyzed the data and helped to draft the manuscript. D.C.S. designed the microarray. D.R. and P.G. helped to design the experimental protocol. M.H. aided in sample collection.

REFERENCES

1. **Allen, M. J., D. C. Schroeder, A. Donkin, K. J. Crawfurd, and W. H. Wilson.** 2006. Genome comparison of two coccolithoviruses. Virol. J. **3:**15.
2. **Allen, M. J., D. C. Schroeder, M. T. Holden, and W. H. Wilson.** 2006. Evolutionary history of the coccolithoviridae. Mol. Biol. Evol. **23:**86–92.
3. **Allen, M. J., D. C. Schroeder, and W. H. Wilson.** 2006. Preliminary characterisation of repeat families in the genome of EhV-86, a giant algal virus that infects the marine microalga *Emiliania huxleyi*. Arch. Virol. **151:**525–535.
4. **Amici, C., A. Rossi, A. Costanzo, S. Ciafre, B. Marinari, M. Balsamo, M. Levrero, and M. G. Santoro.** 2006. Herpes simplex virus disrupts NF-κB regulation by blocking its recruitment on the IκBα promoter and directing the factor on viral genes. J. Biol. Chem. **281:**7110–7117.
5. **Chambers, J., A. Angulo, D. Amaratunga, H. Q. Guo, Y. Jiang, J. S. Wan, A. Bittner, K. Frueh, M. R. Jackson, P. A. Peterson, M. G. Erlander, and P. Ghazal.** 1999. DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. J. Virol. **73:**5757–5766.
6. **Charlson, R. J., J. E. Lovelock, M. O. Andreae, and S. G. Warren.** 1987. Oceanic phytoplankton, atmospheric sulfur, cloud albedo and climate. Nature **326:**655–661.
7. **de Los Santos, T., S. de Avila Botton, R. Weiblen, and M. J. Grubman.** 2006. The leader proteinase of foot-and-mouth disease virus inhibits the induction of beta interferon mRNA and blocks the host innate immune response. J. Virol. **80:**1906–1914.
8. **Edwards, D.** 2003. Non-linear normalization and background correction in one-channel cDNA microarray studies. Bioinformatics **19:**825–833.
9. **Forster, T., D. Roy, and P. Ghazal.** 2003. Experiments using microarray technology: limitations and standard operating procedures. J. Endocrinol. **178:**195–204.
10. **Hancock, D., M. Wilson, G. Velarde, N. Morrison, A. Hayes, H. Hulme, A. J. Wood, K. Nashar, D. B. Kell, and A. Brass.** 2005. maxdLoad2 and maxdBrowse: standards-compliant tools for microarray experimental annotation, data management and dissemination. BMC Bioinformatics **6:**264.
11. **Iyer, L. M., L. Aravind, and E. V. Koonin.** 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. **75:**11720–11734.
12. **Pettus, B. J., C. E. Chalfant, and Y. A. Hannun.** 2002. Ceramide in apoptosis: an overview and current perspectives. Biochim. Biophys. Acta **1585:**114–125.
13. **Schroeder, D. C., J. Oke, M. Hall, G. Malin, and W. H. Wilson.** 2003. Virus succession observed during an *Emiliania huxleyi* bloom. Appl. Environ. Microbiol. **69:**2484–2490.
14. **Schroeder, D. C., J. Oke, G. Malin, and W. H. Wilson.** 2002. Coccolithovirus (Phycodnaviridae): characterisation of a new large dsDNA algal virus that infects Emiliania huxleyi. Arch. Virol. **147:**1685–1698.
15. **Steven, A. C., J. B. Heymann, N. Cheng, B. L. Trus, and J. F. Conway.** 2005. Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. Curr. Opin. Struct. Biol. **15:**227–236.
16. **Suhre, K., S. Audic, and J. M. Claverie.** 2005. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. Proc. Natl. Acad. Sci. USA **102:**14689–14693.
17. **Susin, S. A., N. Zamzami, M. Castedo, E. Daugas, H.-G. Wang, S. Geley, F. Fassy, J. C. Reed, and G. Kroemer.** 1997. The central executioner of apoptosis: multiple connections between protease activation and mitochondria in Fas/APO-1/CD95- and ceramide-induced apoptosis. J. Exp. Med. **186:**25–37.
18. **van Etten, J. L., and R. H. Meints.** 1999. Giant viruses infecting algae. Ann. Rev. Microbiol. **53:**447–494.
19. **Westbroek, P., C. W. Brown, J. Vanbleijswijk, C. Brownlee, G. J. Brummer, M. Conte, J. Egge, E. Fernandez, R. Jordan, M. Knappertsbusch, J. Stefels, M. Veldhuis, P. Vanderwal, and J. Young.** 1993. A model system approach to biological climate forcing—the example of *Emiliania huxleyi*. Global Planet. Change **8:**27–46.
20. **Westbroek, P., J. E. van Hinte, G. J. Brummer, M. Veldhuis, C. Brownlee, J. C. Green, R. Harris, and B. R. Heimdal.** 1994. *Emiliania huxleyi* as a key to biosphere-geoshere interactions, p. 321–334. *In* J. C. Green and B. S. C. Leadbeater (ed.), The haptophyte algae, vol. 51. Clarendon Press, Oxford, United Kingdom.
21. **Westbroek, P., J. R. Young, and K. Linschooten.** 1989. Coccolith production (biomineralization) in the marine alga *Emiliania huxleyi*. J. Protozool. **36:** 368–373.
22. **Wilson, W. H., D. C. Schroeder, M. J. Allen, M. T. G. Holden, J. Parkhill, B. G. Barrell, C. Churcher, N. Hamlin, K. Mungall, H. Norbertczak, M. A. Quail, C. Price, E. Rabbinowitsch, D. Walker, M. Craigon, D. Roy, and P. Ghazal.** 2005. Complete genome sequence and lytic phase transcription profile of a coccolithovirus. Science **309:**1090–1092.